

Predstavitveni portal spletnih jezikovnih virov za slovenščino

**ŠPELA ARHAR HOLDT, KAJA DOBROVOLJC, IZTOK
KOSEM**

Uvod

- S prehodom družbe v digitalno dobo se (tudi) za slovenščino večja število **jezikovnih virov, priročnikov in tehnologij**, ki so za uporabo **prosto dostopni na spletu**.
- Avtorji veliko truda namenjajo zagotavljanju širše dostopnosti in uporabniške prijaznosti izdelkov.
- Vendar za diseminacijo rezultatov v praksi pogosto **zmanjka časa**, izvedba izobraževanj uporabnikov za delo z viri pa tipično presega domet projektov, v katerih se viri pripravljajo.
- Na drugi strani je izziv za uporabnike **raznolikost** spletnih jezikovnih virov, tako v smislu namena oz. vsebine kot konkretnih vmesniških rešitev.
- Tudi če uporabniki določen vir na predstavitvah ali izobraževanjih spoznajo, ob neredni uporabi **izgubijo veščine**, potrebne za uspešno pridobivanje in interpretacijo jezikovnih podatkov.
- Prim. **Jezikovnotehnološko izpopolnjevanje učiteljev v letih 2013 in 2014**

Projekti

- Izdelati **spletno stran**, na kateri so jezikovni viri predstavljeni na **istem mestu** in na **poljuden način**.
- Portal jezikovnih virov: <http://viri.trojina.si>.
- Financirano s pomočjo *Javnega razpisa za sofinanciranje projektov, namenjenih predstavljanju, uveljavljanju in razvoju slovenskega jezika* Ministrstva za kulturo RS.
- Izvedba je potekala v dveh delih (v letih 2014 in 2015): **Izdelava spletne strani z opisi jezikovnih virov in orodij za slovenščino ter osnovnimi (video)navodili za njihovo uporabo** ter **Nadgradnja in popularizacija predstavitvenega portala spletnih jezikovnih virov za slovenščino**.
- Prvo leto smo razvili koncept, pripravili spletno stran in predstavitve za devet jezikovnih virov.
- Drugo leto smo dodali predstavitve za še šest jezikovnih virov in izvedli akcijo *Vir meseca*, v kateri smo na omrežju Facebook predstavljali zanimivosti, povezane z jezikovnimi viri.

PREDSTAVITVENI PORTAL SPLETNIH JEZIKOVNIH VIROV ZA SLOVENŠČINO

Viri Vir meseca O projektu



Dobrodošli na spletni strani, ki predstavlja spletne jezikovne priročnike, vire in storila za slovenščino!

VSI VIRI

jezikovnotehnoški

korpusni

pravopisni

slovarski

slovnčni

terminološki

zgodovinski

vstopna stran

viri, ki so predstavljeni na portalu

o akciji Vir meseca in o projektu

GIGAFIDA

Korpus pisnih besedil Gigafida

Obsežna zbirka sodobnih (1990-2011) slovenskih besedil iz časopisov, revij, knjig, spleta itd., pripravljenih posebej za raziskovanje jezikovne rabe. Korpus obsega skoraj 1,2 milijarde besed.

GOS

Korpus govornjene slovenščine GOS

Posnetki govornjene slovenščine v različnih vsakodnevnih situacijah. Posnetki so transkribirani in umeščeni v zmožljiv vmesnik, s katerim lahko primere govora iščemo, poslušamo in preučujemo.

FRAN

Spletni portal Fran

Na portalu lahko iščemo po različnih slovarjih (splošnih, zgodovinskih, terminoloških, narečnih), svetovalnicah in besedilnih oz. besednih zbirkah Inštituta za slovenski jezik Frana Ramovša.

SLOLEKS

Leksikon besednih oblik Sloleks

Besedne oblike (tako standardne kot nestandardne) za več kot 100.000 najpogostejših slovenskih besed. Za vsako obliko je na voljo podatek o njeni pogostosti in povezava na avtentične primere rabe.

IMP

Jezikovni viri starejše slovenščine

Digitalna knjižnica z več kot 600 knjigami, izdanimi pred letom 1918. Besedila so pripravljena za korpusno raziskovanje, na voljo pa so tudi seznami starejših besed, skupaj s pomenom in primeri rabe.

ŠOLAR

Korpus šolskih besedil Šolar

Korpus prinaša pisna besedila, ki so jih učenci in dijaki slovenskih šol tvorili pri pouku. V precejšnjem delu besedil so posebej označene tudi jezikovne napake, ki so jih v spisih popravili učitelji.

OBELIKS

Oblikoslovni označevalnik Obeliks

Program razdeli slovensko besedilo na besede in vsaki besedi pripiše osnovno obliko, besedno vrsto in dodatne lastnosti (spol, sklon, število itd.). Podatki se izpišejo v tabeli ali v formatu XML.

STAREJŠI
PRAVOPISI**Digitalizirani pravopisi 1899-1962**

Digitalizirane različice petih slovenskih pravopisov (1899, 1920, 1935, 1950, 1962). Pravopisi so za pregledovanje na voljo v zmožljivem spletnem bralniku, za prenos pa tudi v formatu PDF.

STAREJŠE
SLOVNICE**Digitalizirane slovnice 1886-1964**

Digitalizirane različice sedmih slovenskih slovnice (1886, 1900, 1916, 1934, 1940, 1956, 1964). Slovnice so za pregledovanje na voljo v zmožljivem spletnem bralniku, za prenos pa tudi v formatu PDF.

Trenutno so predstavljeni

- 1) Referenčni pisni korpus Gigafida (www.gigafida.net),
- 2) Korpus govornjene slovenščine GOS (www.korpus-gos.net),
- 3) Korpus šolskih besedil Šolar (www.korpus-solar.net),
- 4) Slovarki portal Fran (www.fran.si),
- 5) Slovarki portal Termania (www.termania.net),
- 6) Leksikon besednih oblik Sloleks (<http://www.slovenscina.eu/sloleks>),
- 7) Projekt IMP – viri starejše slovenščine (<http://nl.ijs.si/imp/>)
- 8) Projekt Signor – korpus slovenskega znakovnega jezika (<http://lojze.lugos.si/signor/>),
- 9) Kolaborativni slovar Razvezani jezik (<http://razvezanijezik.org/>),
- 10) Jezikovna svetovalnica ISJFR ZRC SAZU (<http://isjfr.zrc-sazu.si/svetovalnica#v>),
- 11) Terminologišče ISJFR ZRC SAZU (<http://isjfr.zrc-sazu.si/terminologisce#v>),
- 12) Jezikovnodidaktični Pedagoški slovnični portal (www.slovnica.slovenscina.eu),
- 13) Jezikoslovni označevalnik Obeliks (www.slovenscina.eu/tehnologije/oznacevalnik),
- 14) Digitalizirani starejši pravopisi (<http://www.trojina.org/pravopisi>) in
- 15) Digitalizirane starejše slovnice (<http://www.amebis.si/slovnice>).

Posnetki na portalu

- Glavni doprinos Portala jezikovnih virov so **izobraževalni posnetki**, ki na zgoščen, poljuden način **razlagajo vsebino in strukturo** obravnavanega jezikovnega vira,
- in posnetki, ki **na konkretnih primerih uporabe** kažejo, kako lahko v določenem viru najdemo odgovor na specifično jezikovno vprašanje.
- Poleg posnetkov vsako poglavje prinaša tudi povezavo na obravnavani jezikovni vir in projektno stran ter informacije o projektu in avtorjih.
- Prioriteta pri pripravi posnetkov je bila zagotoviti optimalno uporabniško izkušnjo **na različnih vrstah naprav**, od računalnikov z velikimi zasloni do mobilnih telefonov z majhnimi.
- Širšo dostopnost vsebin za različne skupine uporabnikov smo skušali omogočiti tudi s pripravo **slovenskih in angleških podnapisov**.

Posnetki na portalu



PREDSTAVITVENI PORTAL SPLETNIH JEZIKOVNIH VIROV ZA SLOVENŠČINO

Viri

Vir meseca

O projektu



Korpus pisnih besedil Gigafida

Obsežna zbirka sodobnih (1990-2011) slovenskih besedil iz časopisov, revij, knjig, spleta itd., pripravljenih posebej za raziskovanje jezikovne rabe. Korpus obsega skoraj 1,2 milijarde besed.

Povezava na korpus: www.gigafida.net.

Posnetek na desni na kratko predstavlja osnovne značilnosti korpusa Gigafida. Specifični primeri iskanja po korpusu so na voljo v posnetkih spodaj, skupaj z nekaterimi dodatnimi informacijami o projektu in seznamom avtorjev. Priporočamo, da si pred gledanjem posnetkov v nastavitvah predvajalnika vklopite najvišjo razpoložljivo kakovost prikaza (1080p), na voljo pa so tudi slovenski podnapisi.

Osnovne oblike

- Pravi čaj (1993)
- Pravi čaj (14)

Vista besedila

- Časopis (1999)

Vir

- Dnevnik (1150)
- časopis (77)
- Dobrotinski list (71)
- Delo (68)
- Govorenski glasnik (27)
- Več

Leto

- 2002 (62)
- 1999 (57)
- 2001 (52)
- 2000 (49)



kratek opis
in povezava na vir

videoposnetek
(cca. 3 minute)
splošna predstavitev vira

informacije o projektu,
avtorjih in rezultati
akcije Vir meseca

kratki videoposnetki
(cca. 1 minuta)
konkretna jezikovna vprašanja

Primeri iskanja

Več informacij

Avtorji

Vir meseca

Bližnji vzhod, Bližnji Vzhod
ali bližnji Vzhod?

S katerimi samostalniki
se pojavlja beseda **zoprn**?

Dekanja, dekanica
ali dekanika?



Akcija Vir meseca

- V letu 2015 je potekala aktivnost **Vir meseca**, v kateri smo posebej promovirali izbrane vire: **korpus Šolar** (april), **Viri starejše slovenščine IMP** (maj), **korpus Gigafida** (junij), **korpus GOS** (september), **portal Termania** (oktober) in **slovarski iskalnik Fran z Jezikovno svetovalnico in Terminologiščem** (november).
- Promocija je potekala po treh glavnih kanalih: prek omrežja Facebook, v obliki mesečnih e-novičk (*MailChimp*) in na poštnem seznamu SlovLit.
- Tekom akcije smo vsakega od zgoraj navedenih virov preko meseca promovirali z objavami na družbenem omrežju.
- Objave so vsebovale zanimivosti glede priprave vira, informacije o gradivu (npr. različne statistike o jeziku, besedne oblake in druge vrste slikovnega gradiva), predstavitvene posnetke, ideje za uporabo vira v didaktične namene, kratke posnetke pogovora z avtorji in podobno.
- Skupno smo pripravili vsebine za več kot 70 objav, ki so v obliki strnjenih mesečnih poročil prosto na voljo na spletu tudi po zaključku akcije (<http://viri.trojina.si/vir->

Akcija Vir meseca



Akcija Vir meseca

Število ogledov	Osnovna predstavitev	Kratki posnetki	Intervju z avtorjem	Skupaj - vir
Termania	79		69	148
Signor	44			44
Jezikovna svetovalnica	188		94	282
Pedagoški slovnčni portal	125			125
Razvezani jezik	107			107
Terminologišče	94		74	168
Gos	189	129		318
Gigafida	208	171	173	552
Šolar	251	200	211	662
Soleks	248	134		382
IMP	105	149		254
Fran	237			237
Obeliks	128			128
Starejše slovnice	61			61
Starejši pravopisi	78			78
Skupaj - portal				3546

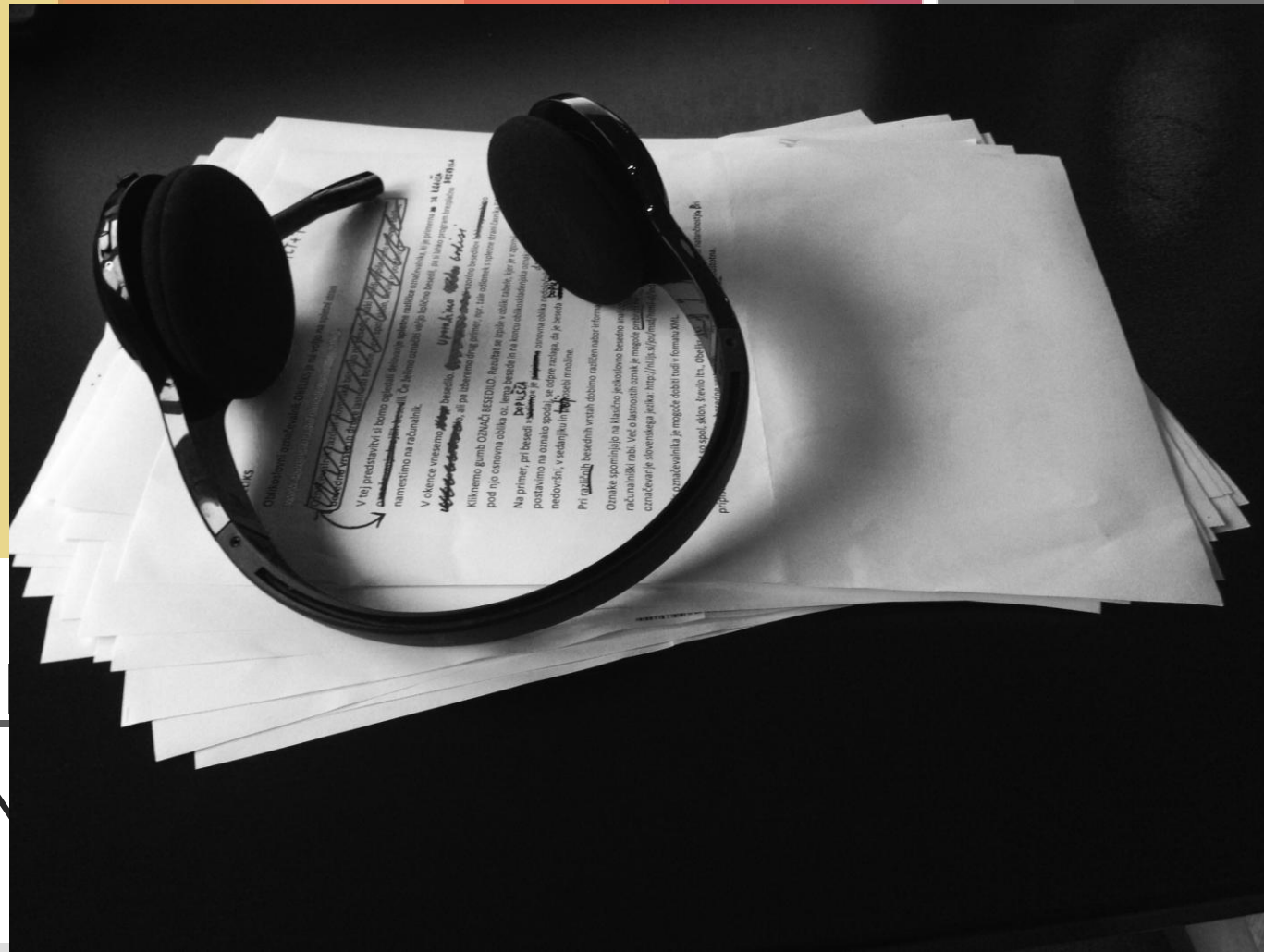
Zaključek

- Portal jezikovnih virov je dosegel velikost, ko ga je mogoče uporabljati kot samostojen vir za (samo)izobraževalne namene.
- Glavna naloga za naprej je dopolniti portal z novimi poglavji in z obstojem vsebin seznaniti čim širši nabor potencialnih uporabnikov.
- Vključiti zainteresirane avtorje, ki bi prek obstoječe platforme lahko ponudili lastne diseminacijske vsebine.
- Dodati rubriko, kjer bi bile sproti predstavljene nadgradnje oz. posodobitve virov, ki se še razvijajo.

Hvala za pozornost!

[HTTP://VIRI.TROJINA.SI](http://viri.trojina.si)
SPELA.ARHAR@TROJINA.SI

trojina



Literatura

- Kozma Ahačič, Nina Ledinek in Andrej Perdih.** 2015. Portal Fran – nastanek in trenutno stanje. V: M. Smolej, ur., *Slovnica in slovar – aktualni jezikovni opis, Obdobja 34*, str. 57–66. Znanstvena založba Filozofske fakultete, Ljubljana.
- Špela Arhar Holdt, Izток Kosem in Polona Gantar.** V pripravi. Corpus-based resources for L1 teaching: The case of Slovene. V: A. Marcus-Quinn, ur.: *Handbook on Digital Learning for K-12 Schools*. Springer.
- Helena Dobrovoljc in Aleksandra Bizjak Končar.** 2015. Pravopisno slovaropisje na Inštitutu za slovenski jezik Frana Ramovša ZRC SAZU. *Slavia Centralis*, 8(1): 34–50.
- Kaja Dobrovoljc, Simon Krek in Tomaž Erjavec.** 2015. Leksikon besednih oblik Sloleks in smernice njegovega razvoja. V: V. Gorjanc, P. Gantar, I. Kosem in S. Krek, ur., *Slovar sodobne slovenščine: problemi in rešitve*, str. 80–105. Znanstvena založba Filozofske fakultete, Ljubljana.
- Kaja Dolar.** 2014. Kolaborativni slovar Razvezani jezik. *Slavistična revija* 62(2): 235–252.
- Tomaž Erjavec.** 2015. The IMP historical Slovene language resources. *Language resources and evaluation* 49/3, str. 753–775.
- Miha Grčar, Simon Krek in Kaja Dobrovoljc.** 2012. Obeliks: statistični oblikoskladenjski označevalnik in lematizator za slovenski jezik V: T. Erjavec, J. Žganec Gros, ur., *Zbornik Osme konference Jezikovne tehnologije, 8. do 12. oktober 2012: zbornik 15. mednarodne multikonference. Informacijska družba – IS 2012, zvezek C*, str. 89–94. Institut Jožef Stefan, Ljubljana.
- Izток Kosem, Mojca Stritar Kučuk, Sara Može, Ana Zwitter Vitez, Špela Arhar Holdt in Tadeja Rozman.** 2012. *Analiza jezikovnih težav učencev: korpusni pristop*. Zavod za uporabno slovenistiko Trojina, Ljubljana.
- Nataša Logar, Miha Grčar, Marko Brakus, Tomaž Erjavec, Špela Arhar Holdt in Simon Krek.** 2012: *Korpusi slovenskega jezika Gigafida, KRES, ccGigafida in ccKRES: gradnja, vsebina, uporaba*. Zavod za uporabno slovenistiko Trojina; Fakulteta za družbene vede, Ljubljana.
- Miro Romih in Simon Krek.** 2012. Termania – prosto dostopni spletni slovarski portal V: T. Erjavec, J. Žganec Gros, ur., *Zbornik Osme konference Jezikovne tehnologije, 8. do 12. oktober 2012: zbornik 15. mednarodne multikonference Informacijska družba - IS 2012, zvezek C*, str. 163–166. Institut Jožef Stefan, Ljubljana.
- Mojca Stritar in Kaja Dobrovoljc.** 2013. Korpusi na poti v šole: jezikovnotehnološko izpopolnjevanje učiteljev. *Slovenščina 2.0*, 1(1): 181–194.
- Darinka Verdonik in Ana Zwitter Vitez.** 2011. *Slovenski govorni korpus Gos*. Ljubljana: Trojina, zavod za uporabno slovenistiko.
- Špela Vintar, Boštjan Jerko in Marjetka Kulovec.** 2012. Compiling the Slovene sign language corpus. V: *8th International Conference on Language Resources and Evaluation, 21–27 May 2012, Istanbul, Turkey. LREC 2012: proceedings*, str. 159–162. ELRA, Istanbul.
- Mojca Žagar Karer.** 2015. Terminologišče – kraj, kjer terminolog išče. *Slavia Centralis*, 8(1): 22–33.